Applying model-based methods to analyze genomics data

by
Zhaohui Qin
Department of Biostatistics, University of Michigan
1420 Washington Heights
Ann Arbor, MI 48109
qin@umich.edu

Abstract

Designing powerful tools to extract scientific insights from massive and noisy high-throughput data is a key challenge in the post-genomic era. Effective clustering and screening algorithms facilitate the identification of subtle patterns of interest, allowing scientists to quickly pinpoint the target of interest for further validation. Model-based approaches offer numerous advantages such as allowing explicit incorporation of biological knowledge and offering reliable statistical inference. We will illustrate the benefits of model-based approaches using examples from DNA sequences and microarray gene expression data analysis. Techniques such as divide-conquer-combine, weighting and feature selection were incorporated to strengthen these model-based approaches.